APPLICATION UNDER UNITED STATES PATENT LAWS

Atty. Dkt. No. PM 265461/PHM70251/UST
(M#)

Invention: ZGGBP1, NOVEL PEPTIDES RELATED TO BIPOLAR AFFECTIVE DISORDERS TYPE 1, SEQUENCES AND USES THEREOF

Inventor (s): FLANNERY, Angela Veronica
FINNEGAN, Maria Christina Martina

Pillsbury Madison & Sutro LLP
Intellectual Property Group

Attorneys
Telephone:

This is a:

- ☐ Provisional Application
- ☐ Regular Utility Application
- ☐ Continuing Application
- ☒ PCT National Phase Application
- ☐ Design Application
- ☐ Reissue Application
- ☐ Plant Application
- ☐ Substitute Specification
  Sub. Spec Filed _____
  in App. No. / _____

SPECIFICATION

ZGGBP1, NOVEL PEPTIDES RELATED TO BIPOLAR AFFECTIVE DISORDER TYPE I, SEQUENCES AND USES THEREOF

This invention relates to a novel human gene (ZGGBP1) associated with affective neurological disorders such as bipolar affective disorder. The invention also relates to homologues of the ZGGBP1 gene in species such as rat and mouse useful in providing animal models of affective disorders. The invention further relates to both the cDNA and the structural gene and to fragments encoding functional domains within the gene. The invention also relates to means for producing the protein encoded by the gene and to means for regulating its production and activity in vivo.

Affective disorders comprise a broad and heterogeneous category of psychiatric illness with a prevalence of up to 20% in the population. The most severe of these disorders is bipolar type I which affects approximately 1% of the population and this rate is fairly consistent across countries. The disease affects young adults, with a mean age of onset of 22 years. Treatment depends upon the phase of the disease and pharmacological agents include lithium carbonate, carbamazepine or valproic acid, tricyclic antidepressants. Monoamine oxidase inhibitors and selective serotonin re-uptake inhibitors are now also being used. The success rate of individual drugs is variable and some patients are treated with a combination of agents, although most have some unwanted side-effects. At present the precise diagnosis of individual affective disorders is difficult and new, gene based, diagnostic methods are desirable.

Family, twin and adoption studies have suggested the importance of genetic predisposition to bipolar affective disorder. On this basis, several groups have undertaken genetic linkage analysis in families with a high incidence of the disorder to find a causal gene. Many of the studies show conflicting data suggesting that a single gene is unlikely to be the cause. Rather, multiple interacting genetic traits may be involved. A recent study (Stine et al. 1995) identified two regions on chromosome 18 showing linkage to the disease.

The present invention is based on our discovery of a novel gene which maps to 18q21 and which unexpectedly shows appreciable sequence homology to the ned-4 gene on chromosome 15. Ned-4 is the human homologue of the mouse nedd-4 gene which is known to be differentially expressed during neural development and to be involved in signal transduction. Human ned-4 has been shown (Schild et al. 1996, Straub

et al. 1996) to be a negative regulator of a sodium channel which is deleted in Liddle's syndrome (a hereditary form of hypertension).

Nedd-4 was originally isolated as a partial cDNA clone from a mouse brain library (Kumar et al. 1992) as one of a set of genes which were differentially expressed during development (Neural precursor cells expressed developmentally down-regulated). The derived amino acid sequence contains three copies of the WW domain (Andre & Springael 1994, Bork & Sudol, 1994; Hofmann & Boucher, 1995), a Ca lipid binding (CaLB/C2) domain (Brose et al. 1995) and a Hect (homologous to the E6-AP carbodyl terminus) domain which has homology to a ubiquitin ligase (E3) enzyme (Huibregtse et al. 1995). The human homologue of nedd-4 (Ned-4) was isolated as an randomly cloned EST (KIAA0093) from immature myeloblast mRNA (Nomura et al. 1994) and shown by sequence comparison to have 86% identity at the amino acid level to the mouse sequence. The human sequence, however, has a fourth copy of the WW domain.

The WW domain is a 40 amino acid sequence found in several unrelated proteins. The two highly conserved tryptophans give it its name. The function of the domain is thought to be involved in protein-protein interactions. Despite their functional diversity, the proteins listed all appear to be involved in cell signalling or regulation. It has been shown that the WW domains of Nedd-4 interact with the proline-rich PY motifs in the epithelial sodium channel in the kidney (Schild et al. 1996). Mutational deletion of the PY motifs in the epithelium sodium channel in Liddle's syndrome, an inherited disease causing systemic hypertension characterised by hyperactivity of the sodium channel, has been shown to abrogate binding of Nedd-4 (Straub et al. 1996). It is therefore likely that Nedd-4 has a negative regulatory role when bound to the channel.

The Hect domain is an E3 ubiquitin-protein ligase domain and enzymes with this domain catalyse polyubiquitination, which is involved in several cellular processes including proteolytic degradation.

The CaLB/C2 domain is thought to be involved in calcium-dependent phospholipid binding, although some proteins containing this domain do not bind calcium and other putative functions for the C2 domain such as binding to inositol -1,3,4,5-tetraphosphate have been suggested. Examples of proteins containing this domain are Protein Kinase C (PKC) isoenzymes and synaptogamins.

PCT patent application WO97/12962 discloses a protein (Pub3) with homology to Pub1, a Schizosaccaromyces Pombe protein which has an apparent function in the ubiquitination of, among other cellular proteins, the mitotic activating tyrosine phosphatase cdc25 and the tumour suppresser protein p53. As such this protein may be involved in regulating the progression of proliferation in eukaryotic cells by effectively controlling the activity of the cdk complexes by modulating the availability of cdc25 and/or p53.

A comparison of Pub3 with ZGGBP1 revealed that the sequences represent two distinct genes which code for two separate, structurally unrelated proteins. The two genes share sequence homology within a certain defined region, the sequences are identical within the region 516-3568 of ZGGBP1, but they do not show any homology within the regions 5' and 3'of this sequence. In addition the derived amino acid sequence for ZGGBP1 is completely different to that derived for Pub 3 as both have been initiated from a different start methionine. A comparison of the nucleotide sequences for ZGGBP1 and Pub 3 is outlined in Figure 5.

Therefore in a first aspect of the present invention we provide the ZGGBP1 gene having the full length cDNA as set out in SEQ ID NO: 1. We further provide fragments of the ZGGBP1 gene comprising ZGGBP1 sequence outside the region defined by base pairs 516-3568 of the ZGGBP1 gene. By fragments we mean contiguous regions of the gene including complementary DNA and RNA sequences, starting with short sequences useful as probes or primers of say about 8-50 bases, such as 10-30 bases or 15-35 bases, to longer sequences of up to 50, 100, 200, 500 or 1000 bases. Indeed any convenient fragment of the gene of say up to 2kb, 3kb, 4kb or more than 4kb may be a useful gene fragment for further research, therapeutic or diagnostic purposes. Further convenient fragments include those whose terminii are defined by restriction sites within the gene of one or more kinds, such as any combination of Rsa1, Alu1 and Hinf1.

In a further aspect of the invention we provide homologues of the ZGGBP1 gene in species such as rat and mouse useful in providing animal models of affective disorders. By homologue, we mean a corresponding ZGGBP1 gene in another species, which displays greater than 85% sequence homology, conveniently greater than 90%, for example 95%, to the human ZGGBP1 sequence. The full sequences of the individual homologues may be determined using conventional techniques such as hybridisation, PCR

and sequencing techniques, starting with any convenient part of the sequence set out in SEQ ID NO: 1. The partial sequence of the mouse gene is set out in SEQ ID NO: 3 and this gene and the protein encoded by this gene represent further independent aspects of the invention.

In a further aspect of the invention we provide polynucleotide sequences capable of specifically hybridising to the ZGGBP1 gene. By specifically hybridising we mean that the polynucleotide hybridises under stringent conditions to the sequence on chromosome 18q21 as set out in SEQ ID No: 1, or to the corresponding non-coding sequence, to the exclusion of other genomic loci. It is contemplated that a species such as a peptide nucleic acid may be an acceptable equivalent to a polynucleotide, at least for purposes that do not require translation into protein.

In a further aspect of the invention we provide a recombinant ZGGBP1 protein obtained by expression of all or a part of the cDNA as set out in SEQ ID NO: 1. The recombinant protein may comprise all or a convenient part of the peptide sequence set out in SEQ ID NO: 2. The production of a protein according to the invention may be achieved using standard recombinant DNA techniques involving the expression of the protein by a host cell as described for example by Sambrook et al. 1989. The isolated nucleic acids described herein may for example be introduced into any convenient expression vector for example the T7 Studier system for expression in E.coli (US-A-4952496), Pichia pastoris for expression in yeast, the Baculovirus system for expression in insect cells and the GS system for expression in mammalian cells by operatively linking the DNA to any necessary expression control elements therein and transforming any suitable prokaryotic or eukaryotic host cell with the vector using well known procedures.

Therefore in a further aspect of the invention we provide a recombinant plasmid comprising all or a part of the ZGGBP1 cDNA of the invention.

The invention further extends to cells containing said recombinant plasmids and to a process for producing a ZGGBP1 protein of the invention which comprises culturing said cells such that the desired protein is expressed and recovering the protein from the culture.

By way of example, the nucleotide sequence in SEQ ID NO: 1 is inserted downstream of the SV40 promoter in the pGEX plasmid vector, and either transiently or

stably expressed in COS -7 cells. Expression of the protein according to the invention can be detected following disruption of the cells by Western blotting .

It may be desirable to produce the individual functional domains of the protein according to the invention in isolation from the rest of the molecule. This may be achieved using the above standard recombination DNA techniques except that in this instance the DNA sequence used is that encoding one of the partial amino acid sequences of the domains identified in Figure 1 or a combination of these.

By way of further example, the nucleotide sequence in SEQ ID NO: 1 is inserted downstream of the SV40 promoter and the glutathione-S-transferase (GST) coding sequence in the pBC plasmid vector, and either transiently or stably expressed in COS -7 cells allowing expression of the corresponding fusion protein. Expression of the fusion protein can be detected following disruption of the cells by Western blotting with antibodies to GST, and furthermore the fusion protein can be used in an affinity binding procedure to find proteins which are functional partners of the protein of the invention from cell extracts.

A ZGGBP1 protein of the invention may in particular be used to screen for compounds which regulate the activity of the enzymes and the invention extends to such a screen and to the use of compounds obtainable therefrom to regulate the activity of the protein in vivo.

Thus according to a further aspect of the invention we provide a method for identifying a compound capable of modulating the action of a ZGGBP1 protein which method comprises subjecting one or more test compounds to a screen comprising (A) a protein containing the amino acid sequence shown in SEQ ID NO: 2 or a homologue or fragment thereof, or (B) the nucleotide sequence shown in SEQ ID NO: 1 or a homologue or fragment thereof, or (C) a host cell expressing a ZGGBP1 polypeptide or a homologue or fragment thereof.

The screen according to the invention may be operated using conventional procedures, for example by bringing the test compound or compounds to be screened and an appropriate substrate into contact with the protein or a cell capable of producing it and determining affinity for the protein in accordance with conventional procedures.

Any compound identified in this way may be used in the treatment of humans and/or other animals of one or more of the above mentioned diseases. The invention thus

extends to a compound selected through its ability to regulate the activity of the protein in vivo as primarily determined in a screening assay utilising the protein containing an amino acid sequence shown in SEQ ID NO: 2 or a homologue or fragment thereof, or a gene coding therefor for use in the treatment of a disease in which the over- or under-activity or unregulated activity of the protein is implicated.

In a further aspect of the invention we provide examples of insertions/deletions and single base change polymorphisms (mutations) as outlined in Figure 6, 7, 8, 9 and 10.

The ZGGBP1 gene of the invention may also be used as the basis for diagnosis, for example to determine expression levels in a human subject, by for example direct DNA sequence comparison or DNA/RNA hybridisation assays. Diagnostic assays may involve the use of nucleic acid amplification technology such as the PCR and in particular the Amplification Refractory Mutation System (ARMS) as claimed in our European Patent No. 0 332 435. Such assays may be used to determine allelic variants of the gene, for example insertions, deletions and/or mutations such as one or more point mutations. Such variants may be heterozygous or homozygous.

In a further aspect of the invention, amplification primers may be provided for use in the above diagnostic methods. In general, these are provided as a set and used for PCR amplification. One of the primers conveniently hybridises to a ZGGBP1 locus outside the region defined by base pairs 516-3568 thus allowing the ZGGBP1 gene on 18q21 to be identified to the exclusion of other loci.

The ZGGBP1 gene may also be used in gene therapy, for example where it is desired to modify the production of the protein in vivo, and the invention extends to such uses.

Knowledge of the gene according to the invention also provides the ability to regulate its expression in vivo by for example the use of antisense DNA or RNA. Thus, according to a further aspect of the invention we provide an antisense DNA or an antisense RNA which is complementary to the polynucleotide sequence shown in SEQ ID NO: 1. By complementary we mean that the two molecules can base pair to form a double stranded molecule.

The antisense DNA or RNA for co-operation with the gene in SEQ ID NO: 1 can be produced using conventional means, by standard molecular biology and/or by chemical synthesis as described above. If desired, the antisense DNA or antisense RNA may be

chemically modified so as to prevent degradation in vivo or to facilitate passage through a cell membrane and/or a substance capable of inactivating mRNA, for example ribozyme, may be linked thereto and the invention extends to such constructs.

The antisense DNA or antisense RNA may be of use in the treatment of diseases or disorders in humans in which the over- or under-regulated production of the gene product has been implicated. Such diseases or disorders may include those described under the general headings of neurologic, eg.stroke, dementia, renal eg. hypertension, nephrosis, cardiovascular disorders.

Convenient DNA sequences may be obtained using conventional molecular biology procedures, for example by probing a human genomic or cDNA library with one or more labelled oligonucleotide probes containing 10 or more contiguous nucleotides designed using the nucleotide sequences described here. Alternatively, pairs of oligonucleotides one of which is homologous to the sense strand and one to the antisense strand, designed using the nucleotide sequences described here to flank a specific region of DNA may be used to amplify that DNA from a cDNA library.

The ZGGBP1 protein of the invention and homologues or fragments thereof may be used to generate substances which selectively bind to it and in so doing regulate the activity of the protein. Such substances include, for example, antibodies, and the invention extends in particular to an antibody which is capable of recognising one or more epitopes containing the protein binding domains shown in Figure 1. In particular the antibody may be neutralising antibody.

As used herein the term antibody is to be understood to mean a whole antibody or a fragment thereof, for example a F(ab)2, Fab, FV,. VH or VK fragment, a single chain antibody, a multimeric monospecific antibody or fragment thereof, or a bi- or multi-specific antibody or fragment thereof.

The invention will now be illustrated but not limited by reference to the following detailed description, References, Examples and Figures wherein:

Figure 1 shows the predicted amino acid sequence of ZGGBP1. The C2 domain is indicated by carets , the four WW domains are indicated by asterisks and the Hect domain is indicated by underlining .

Figure 2 shows a comparison of amino acid sequences of human ned4 Swissprot entry P46934 and ZGGBP1.

Figure 3 shows a Northern blot analysis of various human tissues probed with ZGGBP1.

Figure 4 shows a comparison of the nucleic acid sequences of human and mouse ZZGBP1. The mouse sequence is a partial cDNA which spans the C-terminal portion of the human protein coding region.

Figure 5 shows a comparison of the nucleic acid sequences for ZGGBP1 and Pub3

Figure 6 shows a polymorphism located at position 3554 of the cDNA sequence

Figure 7 shows a polymorphism located at position 4828 of the cDNA sequence

Figure 8 shows a polymorphism located in an intronic sequence derived from a BAC containing ZGGBP1

Figure 9 shows a variable number of tetranucleotide repeats located within an intronic sequence from ZGGBP1

Figure 10 shows an insertion at position 4032 of the cDNA sequence

References

Stine O.C. et al. (1994) Am.J.Hum.Genet. *57*, 1384-1394;

Lovett M., Kere J. and Hinton L. (1991) Proc.Natl.Acad.Sci. USA *88*, 9628-9632;

Schuler G.D. et al. (1996) Science *274* , 540-546;

Kumar S., Tomooka Y. and Noda M. (1992) Biochem.Biophys.Res.Commun. *185*, 1155-1161;

Andre B. and Springael J.Y. (1994) Biochem.Biophys.Res.Commun. *205*, 1201-1205;

Bork P. and Sudol M. (1994) Trends Biochem.Sci. *19*, 531-533;

Hofmann K. and Bouchet P. (1995) FEBS Letts. *358*, 153-157;

Brose N., Hofmann K.O., Hata Y., Suedhof T.C.(1995) J. Biol. Chem. *270*, 25273-25280;

Huibregtse J.M., Scheffner M., Beaudenon S. and Howley P.M. (1995) Proc.Natl.Acad.Sci. USA, *92*, 2563-2567;

Nomura N. et al. (1994) DNA Res. *2*, 37-43;

Schild L., Lu Y., Gautschi I., Schneeberger E., Lifton R.P. and Rossier B.C. (1996) EMBO J. *15*, 2381-2387;

Straub O., Dho S., Henry P.C., Correa J., Ishikawa T., McGlade J. and Rotin D. (1996) EMBO J. *15* , 2371-2380;

Shizua H., Birren B., Kim U-J, Mancino V., Slepak T., Tachiiri Y. and Simon M. (1992) Proc.Natl.Acad.Sci. USA *89*, 8794-8797;

Kim U-J., Shizuya H., Kang H-L., Choi S-S., Garrett C.L., Smink L.J., Birren B.W., Korenberg J.R., Dunham I. and Simon M. (1996) Proc.Natl.Acad.Sci. USA *93*, 6297-6301.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) "Molecular Cloning: A Laboratory Manual (2nd edition)" Cold Spring Harbor Laboratory Press NY

Example 1

Identification of ZGGBP1

We used two methods for investigating the 18q21 region of interest. In one method we used positional cloning to identify novel transcripts from physical clones representing the region and in a second method we utilised public databases to identify transcripts which had been assigned to a low resolution map of the region by radiation hybrid mapping and assigned them to physical clones representing a high resolution map of the region.

Method 1 - Positional Cloning

The 18q21 region described by Stine et al. (1995) is delimited by the STS markers used by that group to identify linkage. They found the most strongly linked marker to be D18S41, which had a LOD score of 3.51 in cases of paternal inheritance. Linkage declined over flanking markers. We identified a set of four Yeast Artificial Chromosomes (YACs) which comprised a contiguous overlapping set of genomic clones covering the defined region by the presence in those YACs of STS markers used in the Stine study.

DNA from the YACs was prepared and used in a PCR-based hybridisation approach to enrich for transcripts from a human fetal brain cDNA library. This approach, known as direct selection (Lovett et al. 1991) has been shown to be efficient in identifying transcripts present on large genomic clones.

Method 2 - Refining Radiation Hybrid Mapped Transcripts

The UNIGENE database is a repository for transcripts which have been mapped by taking representative Expressed Sequence Tagged Sites (ESTs) and performing PCR analysis on a panel of radiation hybrids which have been calibrated with respect to a framework of 1000 genetic markers (Schuler et al. 1996). We found 36 EST clusters which had been mapped to a radiation hybrid map interval which corresponded to the 18q21 region of interest and to flanking regions outside.

All the ESTs were tested by PCR on our YAC genomic clones to determine which were present. We found approximately half of the ESTs to be present within the genomic clones and were able to order them based on their position within the YAC contig.

Results

Several clones from our direct selection experiments showed sequence homology to a known EST which we had previously shown to be present in two of the YACs within the contig. The EST was representative of a cluster of sequences. All of these sequences were assembled together using DNAStar Seqman and the consensus sequences obtained were used iteratively to search for other database members within both Unigene, dbEST and EMBL databases. This resulted in the surprising identification of two further clusters of ESTs which had previously not been related to each other on the basis of sequence analysis. The two new EST clusters were annotated as having sequence similarity to ned-4. This was an unexpected finding since we had recently mapped the human ned-4 by Fluorescence In Situ Hybridisation (FISH) to chromosome 15. We were aware that ned-4 was involved in neuronal cell signalling and we concluded that the EST cluster on 18q21 must represent a closely related gene and therefore likely to be involved in affective neurological disorders such as bipolar affective disorder.

The assembly of the EST clusters did not give rise to a single complete contiguous sequence. The reason for this is that many of the EST sequences were derived from IMAGE cDNA clones for which end sequence only was available. In order to fill in the gaps and give a complete contig, four of these clones (IMAGE I.D. 80951, 33059, 79526 and 79984) were sequenced completely to fill the gaps and give an entire complete contiguous sequence. Comparison of the sequence with ned-4 showed that the contig comprised 2kb of 3'Untranslated Region (UTR) and 700bp of the coding region of a gene

which had approximately 85% identity at the amino acid level to ned-4 and which we named ZGGBP1.

Isolation of the full length gene for ZGGBP1

The extending of partial transcripts to full length clones can be a complex and difficult process requiring skill and expertise for success. Having considered several possibilities, we opted for a PCR-based approach to isolate and characterise the full length ZGGBP1 gene. Human foetal brain double stranded cDNA was synthesised from mRNA using standard methods (Sambrook et al 1989) and ligated into lambda Zap vector by use of adapters. However, in order to minimise the loss of transcripts often seen following the cloning step, the resulting ligation mix was not cloned but was instead used as a template for PCR. Oligonucleotide primers specific to ZGGBP1 were used in combination with vector specific primers to amplify DNA across the unknown part of the gene. Since the distance to be covered was unknown, we performed long PCR using the commercially available BCL Expand enzyme and long (30mer) oligonucleotide primers. Since we were using unamplified material, where our target cDNAs were likely to be present only in very small amounts, we utilised a secondary PCR step with nested oligonucleotide primers and again using long PCR to yield sufficient PCR products to be visible by gel analysis and also to minimise the possibility of non-specific PCR amplification. The PCR products derived from these experiments were then purified and sequenced directly. Where necessary, the DNA sequence obtained was used to design further primers to walk along the gene in a 3' - 5' direction. The complete nucleotide sequence derived from this work is 5.2kb and the translated amino acid sequence is shown in SEQ ID NO: 1.

The amino acid sequence derived from the cDNA was compared with that of ned-4 and is shown in Figure 2. The proteins diverge markedly towards the N-terminal portion of the protein, although there is conservation of the common functional motifs.

Northern analysis using a probe derived from the 3'UTR of ZGGBP1 showed a band at approximately 4.8kb but also a more abundant band of 9kb in size in several neurological tissues, with the exception of medulla or spinal cord. These bands are likely to be due to alternative splicing (Figure 3). Other tissues contained the 4.8kb band at higher abundance with respect to the 9kb band and also a 4kb band. ZGGBP1 was

expressed in all tissues examined with the exception of liver where we could not detect a transcript at our current detection sensitivity.

Comparison of Amino Acid Sequences of human ned-4 and ZGGBP1

A comparison of the amino acid sequences of human ned-4 and ZGGBP1 is shown in Figure 6. The two proteins have a high level of homology over much of the C-terminal region, including the Hect and WW domains, but diverge over the central portion of the protein. There is a further block of homology near to the N-terminal region, including the C2 domain. The presence of these domains in ZGGBP1 suggests some common functionality with ned-4.

Identification of polymorphic variants of ZGGBP1

500bp regions of the ZGGBP1 cDNA were PCR amplified from a variety of tissues and lymphoblastoid cell lines. Sequencing was carried out and polymorphisms identified as outlined in Figures 5 and 6. Some intronic sequence had been identified from a genomic clone and sequence analysis of these regions identified a further polymorphic variant as outlined in Figure 7. A tetranucleotide repeat (GATT) was also identified in an intronic sequence derived from this BAC and this was found to have variable numbers of repeats (Figure 8).

Isolation of Genomic Clone for ZGGBP1

The Research Genetics human Bacterial Artificial Chromosome (BAC) library (Shizua et al. 1992, Kim et al. 1996) was screened by PCR using primers specific to the 3'UTR of ZGGBP1 and BACs were isolated. These are being used to characterise the structural gene including the intron/exon structure and the 5' regulatory region.

Isolation of Mouse homologue for ZGGBP1

The full length sequence of ZGGBP1 shown in SEQ ID NO: 1 was used to search the dbEST database to identify homologous mouse sequences. Three overlapping IMAGE clones were identified (IMAGE I.D.479436, 573510, 482922) comprising a partial transcript. Comparison of the mouse and human nucleotide sequence is shown in Figure 4. The mouse clones were isolated for use as a probe for in situ hybridisation on sections

of mouse brain during development, and as a probe of mouse genomic libraries to isolate genomic clones and to produce transgenic mice by gene targeting using homologous recombination.